При составлении тестовых заданий по КСЕ учитывались все возможности и ограничения программного комплекса, предусмотренные разработчиком.

Система позволяет создавать тестовые задания закрытого типа с простым и множественным выбором ответа, открытого типа (ввод ответа с клавиатуры).

Тесты простого выбора предусматривают выбор одного из предложенных вариантов. Данный вид тестовых заданий интуитивно понятен обучающимся, требует минимального времени для ввода ответа, прост в обработке ответов, однако обеспечивает большую вероятность угадывания ответа.

Тесты множественного выбора предполагают выбор нескольких вариантов из предложенных ответов. Данный вид заданий обеспечивает высокую степень информативности, позволяет учесть частично правильные ответы, однако существует вероятность угадывания правильного ответа, отсутствует общепризнанная процедура обработки ответов.

При выполнении тестовых заданий открытого типа обучающемуся предлагается самостоятельно сформулировать и ввести ответ с клавиатуры. В данном случае вероятность угадывания минимальна, однако имеется сложность, связанная с синтаксическим анализом ответов. Указанный тип заданий эффективен при проверке терминов, дат, констант.

С учетом указанных преимуществ и недостатков разных видов заданий, а также в зависимости от темы, сложности рассматриваемых вопросов и понятий используются тестовые задания различных видов.

Разработанные тестовые задания апробированы при обучении студентов филологических специальностей педагогического вуза дисциплине «Концепции современного естествознания» в течение нескольких семестров, а также при подготовке к Интернет-экзамену в сфере высшего профессионального образования (апрель 2006 г.) по дисциплине «Концепции современного естествознания» для студентов по ООП 033200 и 032900 Российского государственного педагогического университета им. А.И.Герцена.

Как показывает опыт, создание теста, соответствующего всем требованиям, и имеющего устойчивые коэффициенты надежности и валидности, представляет определенную трудность. Вместе с тем, тестирование при обучении дисциплине «Концепции современного естествознания» играет очень важную роль и может рассматриваться как важнейшее средство мониторинга учебных достижений студентов.

Кромер В.В.

О ТОЧНОСТИ ТЕСТОВЫХ ИЗМЕРЕНИЙ

kromer@newmail.ru
Русско-немецкий университет
г. Новосибирск

Тестологу В.С. Аванесову принадлежит определение теста как шкалы. Определение очень образное, латинское слово *scala* означает *лестница, ступеньки*. Сколько-же ступенек в этой лестнице? Первый напрашивающийся ответ – столько, сколько в тесте заданий – неверен. Под высотой ступеньки в данном случае следует понимать значимо определяемую тестом разницу между уровнями измеряемого свойства. Исходя из данного определения, количество различаемых уровней (ступенек) определится как диапазон теста (разница между предельными значениями измеряемого тестом свойства), деленный на "высоту ступеньки". В рамках классической тестовой теории удобно проводить оценку в единицах z-шкалы – стандартной шкалы с нулевым средним значением и единичным стандартным отклонением. Само значение оценки уровня подготовленности испытуемого, выраженное в единицах z-шкалы, является случайной переменной, распределенной по нормальному закону со стандартным распределением s_e, зависящим от r – надежности применяемого теста: $s_e = s_y\sqrt{1-r}$, где s_y – стандартное отклонение тестовых результатов. Для z-шкалы $s_y = 1$, и $s_e = \sqrt{1-r}$. При известных лимитах теста $z_{\max}$ и $z_{\min}$ диапазон измерения теста составляет $(z_{\max} - z_{\min})$, и для определения количества градаций (ступенек) теста встает задача определения высоты "ступеньки" в единицах z-шкалы. Рациональное задание этого значения требует возврата к вопросу о форме распределения тестовых результатов.

Форма данного распределения постулируется исходя из интуитивных соображений и проверяется практикой. Подробное рассмотрение данного вопроса выходит за рамки данного доклада, в классической тестовой теории за стандарт принимается нормальное распределение. Известно, что "хвосты" нормального распределения короче, чем у более длиннохвостовых логнормального и Коши распределений, однако на практике и оно оказывается неудобным именно ввиду наличия хвостов, которым не находится соответствие в репрезентативной нормативной выборке, а тем самым и в генеральной совокупности испытуемых (даже бесконечной по объему, поскольку измерение производится тестом – инструментом с конечными пределами измерения). Возможно использование усеченного нормального распределения, однако существует и иной путь стандартизации формы конечного распределения.

В данном докладе за стандарт распределения предлагается принять сумму из n равномерных (прямоугольных) распределений. При $n=1$ имеем исходное прямоугольное распределение, $n=2$ соответствует треугольному распределению, $n \geq 3$ – распределению колоколообразной формы, с ростом n все более неотличимым от предельного (при $n \to \infty$) нормального распределения. Значение n выбирается исходя из наших представлений о размахе распределения, который равен $\sqrt{12n}$ в единицах z-шкалы. Ожидаемый размах определяется размером выборки и табулирован.

За стандарт распределения тестовых баллов предлагается принять сумму трех прямоугольных распределений, что отвечает реальным размерам выборок и известному правилу "трех сигм", поскольку при $n=3$ размах равен 6 стандартным отклонениям.

При разделении диапазона измерений теста на конечное число градаций измеренное значение тестового результата может относиться к любой из градаций (если измеренное значение нормально распределено), но с разной вероятностью. Теорема Байеса позволяет определить эти вероятности, а фишеровский принцип максимального правдоподобия отнести это значение к наиболее вероятной градации (ближайшей ввиду симметричности распределения). Приняв расстояние между соседними градациями за Δ, вычислим вероятность измеренного значения x принадлежать дискретной отметке z_m введенной шкалы:

$$p_{z_m} = \frac{f\left(\frac{x - z_m}{s_e}\right)}{f\left(\frac{x - z_m}{s_e}\right) + f\left(\frac{x - z_{m+1}}{s_e}\right) + \ldots + f\left(\frac{x - z_{m-1}}{s_e}\right) + \ldots}$$, где f(z) – функция плотности принятого распределения, а в знаменателе формулы – сумма плотностей вероятностей функции распределения для всех распределений, сдвинутых относительно z_m на $\pm i\,\Delta$, где i – любое целое число. Чтобы распределение этой вероятности повторяло распределение вероятности ошибки, достаточно постоянства знаменателя в вышеприведенной формуле. При постулировании распределения тестовых результатов в соответствии с предлагаемым распределением (с некотором n) целесообразна замена нормального распределения ошибки измерения на предлагаемое распределение с тем же n. Тогда при $\Delta = \frac{s_e\sqrt{12n}}{n} = s_e\sqrt{\frac{12}{n}}$ сумма плотностей предлагаемых распределений с параметром n совершенно равномерно заполнит ось абсцисс. Количество градаций определяется как $k = \frac{\sqrt{12n}}{\Delta} = \frac{n\sqrt{12n}}{s_e\sqrt{12n}} = \frac{n}{s_e} = \frac{n}{\sqrt{1-r}}$.

Для вполне достаточного для практики значения $n = 3$ количество градаций изменяется от 5 (для тестов сомнительной надежности с $r = 0{,}64$ до 12 (для тестов с отличной надежностью 0,94). Для тестов с удовлетворительной надежностью 0,75 (самых распространенных) количество градаций составляет 6. Таким образом, количество градаций "учительских" (home-made) тестов равно количеству градаций шкалы школьных баллов (при введении в оборот "заслуженной двойки" и отграничении ее от "единицы").

При объединении с равным весом результатов тестирования несколькими тестами число градаций суммарной оценки возрастает как корень квадратный из количества тестов. Таким образом, шкала со 100 градациями начинает реально работать при свертке результатов около сотни независимых измерений, что на практике реализуется, например, в условиях вычисления рейтингового показателя выпускника вуза.

Предложенная концепция оценки количества градаций тестового результата (т.е. разрешающей способности теста) имеет аналогию в физическом мире. В теории оптических приборов пределом разрешения называется наименьшее угловое или линейное расстояние между двумя точками, начиная с которого их изображения сливаются. Аналогией распределения ошибки является распределение освещенности по сечению размытого в результате дифракции или оптических аберраций пятна (изображения точки), а выдвинутое условие постоянства знаменателя в формуле Байеса соответствует требованию равномерной освещенности поля изображения, т.е. слиянию изображений отдельных точек.